SAMPLING ERROR ESTIMATION IN THE 1995 CENSUS TEST FOR SMALL AREAS

Thomas R. Krenzke and Alfredo Navarro[1]
Thomas R. Krenzke, U.S. Census Bureau, Washington, D.C. 20233

KEY WORDS: Nonresponse Followup, Integrated Coverage Measurement

1.0 Introduction

In Census 2000, sampling techniques will be implemented for two purposes. The integrated coverage measurement (ICM) program is needed to resolve the undercount in the census numbers. The nonresponse followup (NRFU) program will consist of a sample of housing units (HUs) that did not return their questionnaires. This is in contrast with traditional census methods of following-up on all non-mail returns and is expected to be more cost efficient. Estimating the error due to sampling for any published estimate is a policy of the Census Bureau. In addition, the Census Bureau has legal requirements to publish census numbers for blocks for use in congressional redistricting. The block-level data have to be cross-tabulated by the following redistricting categories: race (5 categories), Hispanic origin (2), age (2), and total population. There are 39 total data items for which estimates at the block-level are needed. This introduces the problem of estimating sampling error for small areas. The challenge is to not only compute variances for the small areas, but also to compute variances for large areas not yet defined like congressional districts. A brief overview of the sample design and the estimation methods is given in section 2. The procedures for measuring sampling errors associated with census numbers for small areas from the 1995 Census Test are explained in section 3. The assumptions and limitations of the methodology used to estimate the error due to ICM and NRFU sampling are discussed. Comparisons of variance generalization methods are discussed and the resulting procedures are presented in section 4. The interested reader may contact the authors for a longer version which includes a more detailed discussion.

2.0 Sample Design and Estimation Methods

For the 1995 Census Test, the ICM sample consisted of a systematic sample of block clusters. Blocks were grouped into clusters of 30 or more HUs and then stratified into ICM sampling strata defined by racial and ethnic composition and size of the cluster. Further stratification defined groups of blocks. Each stratification group was split into two panels, then a systematic sample of block clusters was selected within ICM sampling stratum and panel. In ICM selected blocks, all nonresponding HUs were in the NRFU sample. In non-ICM selected blocks, a stratified systematic sample was selected from the nonresponse universe. The NRFU sample explored two sampling strategies: block sampling where the block cluster was the ultimate sampling unit, and unit sampling where the HU was the ultimate sampling unit. Block sampling was implemented in all three 1995 Census Test sites (Oakland, CA, Paterson, NJ, and NW Louisiana). Due to budget constraints, unit sampling was done only in Oakland.

The ICM poststrata were defined with the intent of producing direct estimates of population for various demographic domains for each site. The categories for the poststrata are race/ethnicity (4), tenure (2), and age/sex (7). There was a different number of poststrata for each test site (Oakland (56), Paterson (42), and NW Louisiana (28)) because some poststrata were collapsed due to the small numbers in certain categories. The post-NRFU estimator in the 1995 Census Test for block i and ICM poststratum k is $\hat{C}_{ki} = SR_{ki} + \hat{NR}_{ki}$, where SR_{ki} is the person count from census self responses in block i for poststratum k, and $\hat{NR}_{ki}$ is a weighted estimate of persons in census nonresponse units in block i for poststratum k.

Two estimation methods were used in the ICM program to resolve the undercount, Dual System Estimation (DSE) and CensusPlus. Both DSE and CensusPlus collected data from census questionnaires in the ICM operations, but the estimation procedures for the calculation of the ICM poststratum factors, $\hat{F}_k$, differed somewhat. See Killion (1996) for details on the ICM operation. Both DSE and CensusPlus were used for the Oakland and Paterson sites, but only CensusPlus data was used for NW Louisiana. The variances of poststratum DSEs in Oakland and Paterson are not expected to differ from the corresponding variances from CensusPlus estimates (Schindler and Navarro, 1994).

3.0 Statistical Methodology for Direct Variances

This section outlines the procedure to be implemented for the calculations of block and tract-level variances for the redistricting data item estimates in the 1995 Census Test. For redistricting data item K for geographic domain I, the final estimate is an aggregate of ratio estimates of the total population for block i and poststratum k,

$$\hat{B}_{KI} = \sum_{k \in K} \sum_{i \in I} \hat{F}_k \hat{C}_{ki}$$

where, $\hat{F}_k = \hat{Y}_k / \hat{C}_k$, $\hat{Y}_k$ is the estimate from the ICM sample for poststratum k, and $\hat{C}_k$ is the estimate of the census total for poststratum k based only on ICM selected blocks. A

limitation of the synthetic estimator, $\hat{B}_{kI}$, is that within ICM poststratum k, the census coverage rate is not allowed to vary from block to block. This presents a bias in the estimation procedure. The variances presented below do not reflect this bias in the synthetic estimator (see Bell, 1996). More research is needed in order to provide estimates that allow for variation in census coverage over blocks. The variance of the estimate, $\hat{B}_{kI}$, can be obtained from the decomposition which conditions on the ICM sample, s_1. The NRFU sample is denoted by s_2.

This decomposition partitions the total variance into two components:

$$Var(\hat{B}_{kI}) = \underset{s_1\ s_2}{Var}[E(\hat{B}_{kI}|s_1)] + \underset{s_1\ s_2}{E}[Var(\hat{B}_{kI}|s_1)]$$

where, $E_{s2}(\hat{B}_{kI}|s_1)$ denotes the average over all possible NRFU samples given the ICM sample, and the first term then is the variance of $E_{s2}(\hat{B}_{kI}|s_1)$ over all possible ICM samples, which is the sampling error due to ICM sample. The $Var_{s2}(\hat{B}_{kI}|s_1)$ is the conditional variance over all possible NRFU samples given the ICM sample, and $E_{s1}[\ Var_{s2}(\hat{B}_{kI}|s_1)]$ is the average over all ICM samples of the $Var_{s2}(\hat{B}_{kI}|s_1)$. This second term is the sampling error due to NRFU sampling. For the ICM variance component of the total variance,

$$\begin{aligned}
\underset{s_1\ s_2}{Var}[E(\hat{B}_{kI}|s_1)] &= \underset{s_1\ s_2}{Var}[E(\Sigma\Sigma\hat{F}_k\hat{C}_{ki}|s_1)] \\
&= \underset{s_1}{Var}[\underset{k}{\Sigma}\hat{F}_k E(\underset{i\in I}{\Sigma}\hat{C}_{ki}|s_1)] \\
&= \underset{s_1}{Var}[\underset{k}{\Sigma}\hat{F}_k E(\hat{C}_{kI}|s_1)]
\end{aligned}$$

since $\hat{F}_k$ is fixed given the ICM sample. Suppose we assume that $\hat{C}_{kI}$ is an unbiased estimator of $C_{kI} = \Sigma_{i\in I}C_{ki}$, the census total for poststratum k and geographic domain I that would be obtained under 100% followup (Note: $\hat{C}_{ki}=C_{ki}$ in ICM blocks and blocks in the NRFU block samples). Then, $E(\hat{C}_{kI}|s_1)=C_{kI}$, and

$$\begin{aligned}
\underset{s_1\ s_2}{Var}[E(\hat{B}_{kI}|s_1)] &\approx \underset{s_1}{Var}(\underset{k}{\Sigma}\hat{F}_k C_{kI}) \\
&= \underset{k\in K I\in K}{\Sigma\Sigma} cov(\hat{F}_k,\hat{F}_l)(C_{kI})(C_{lI})
\end{aligned}$$

The variances and covariances of the ICM poststratum factors, $\hat{F}_k$, were calculated using the jackknife method by Town and Fay (1995). It should also be noted that before calculating the block-level ICM variances, the block-level data records were combined until the 'cluster of blocks' had a total population count greater than 50 (for block-level variances only). This was done so that more reliable 'block-level' variances were used when generalizing the variances (Section 4.0). The above equation estimates the error due to variance in the estimated ICM poststratum factors. Therefore, the second variance component, $E_{s1}[\ Var_{s2}(\hat{B}_{kI}|s_1)]$, which is error due to NRFU sampling, is needed to measure total sampling error in the 1995 Census Test.

For the NRFU component of the total variance,

$$\begin{aligned}
\underset{s_1\ s_2}{E}(Var(\hat{B}_{kI}|s_1)) &= \underset{s_1\ s_2}{E}(Var(\underset{k}{\Sigma}\hat{F}_k\hat{C}_{kI}|s_1)) \\
&= \underset{s_1}{E}(\underset{k\ l}{\Sigma\Sigma}\hat{F}_k\hat{F}_l\ \underset{s_2}{Cov}(\hat{C}_{kI},\hat{C}_{lI}|s_1)) \\
&= \underset{k\ l}{\Sigma\Sigma}\underset{s_2}{Cov}(\hat{C}_{kI},\hat{C}_{lI})\ \underset{s_1}{E}(\hat{F}_k\hat{F}_l) \\
&= \underset{k\ l}{\Sigma\Sigma}\underset{s_2}{Cov}(\hat{C}_{kI},\hat{C}_{lI})\ [\underset{s_1}{E}(\underset{s_1}{Cov}(\hat{F}_k,\hat{F}_l) \\
&\qquad + \underset{s_1}{E}(\hat{F}_k)\ \underset{s_1}{E}(\hat{F}_l)]
\end{aligned}$$

In order to make computations easier, the procedure to approximate the variance due to NRFU sampling assumes unit sampling within each site, simple random sampling, assumes that $E(\hat{F}_k)=1$, $Cov(\hat{F}_k,\hat{F}_l)=0$ for all k,l, and $Cov(\hat{C}_{ki},\hat{C}_{li})=0$ for $k\neq l$ and do not depend on s_1. More research is needed to address the sensitivity of the variance estimates to these assumptions. With those assumptions, we have,

$$\underset{s_1\ s_2}{E}(Var(\hat{B}_{kI}|s_1)) \approx \underset{k}{\Sigma}\ Var(\hat{C}_{kI}) \approx Var(\underset{k}{\Sigma}\hat{C}_{kI})$$

The NRFU sampling error component of the total variance is approximated by a design-based variance estimator,

$$\underset{s_1\ s_2}{E}[Var(\hat{B}_{kI}|s_1)] \approx \frac{\sum_{g=1}^{2}\sum_{h} M_{ghI}(M_{ghI} - m_{ghI})\ s^2_{Kgh}}{m_{ghI}}$$

where, m_{ghI} is the number of sampled nonrespondents for panel g, ICM stratum h, and geographic domain I, and M_{ghI} is the total number of non-mail returns in panel g, ICM stratum h, and geographic domain I. For blocks associated with block sampling, we simulate unit sampling by letting $m_{ghI} = f\,M_{ghI}$, where f is the sampling fraction equal to 1/(3.5) in Oakland and 1/6 in Paterson and NW Louisiana. The estimated population variance of the number of persons per HU for redistricting item K, panel g, ICM sampling stratum h, is,

$$s^2_{Kgh} = \frac{\sum_{j=1}^{m_{Kgh}} (C_{Kghj} - \bar{C}_{Kgh})^2}{m_{Kgh} - 1}$$

where, C_{Kghj} is the number of persons in HU j, for redistricting item K, panel g, and ICM stratum h, $\bar{C}_{Kgh}$ is the average number of persons/HU for redistricting item K, panel g, and ICM stratum h. There were many block clusters with very few sampled units, so to estimate S^2 at the poststratum level would result in unreliable variances. Therefore, S^2 was estimated at a higher aggregated level, which was the ICM sampling strata, and was assumed to hold at the geographic domain of interest.

The standard errors due to ICM sampling, relative to the estimated total, range from about 2% to 6%. These relative standard errors (RSE), or coefficients of variation, do not seem to be highly related to the magnitude of the estimated totals. The reason for this is that the ICM RSE of the estimated block-level total for poststrata k, is theoretically equal to the RSE of the ICM site-level factors for poststrata k. That is, within poststratum k, the RSE is constant for all geographic domains. When the NRFU variance component is added to the ICM variance component, the resulting relative variance is related to the estimated total. The same NRFU sampling fraction is applied for all demographic groups, so for people of races and ethnic background that are not numerous, the estimated number of people will have large RSEs. As shown in Table 1, at the block cluster level, NRFU sampling error component is the dominating component.

Table 1. Median % Contribution--NRFU Component
Oakland Block Cluster Level (Tract Level)

Race and Ethnicity	All Persons	Age <18	Age 18+
All persons	94 (41)	99 (83)	91 (32)
Hispanic origin	99 (93)	100 (96)	99 (92)
White	100 (96)	100 (97)	100 (95)
Black	100 (99)	100 (100)	100 (99)
AI,Eskimo,Aleut	100 (100)	100 (100)	100 (99)
Asian, PI	100 (99)	100 (100)	100 (99)
Other race	100 (97)	100 (99)	100 (97)
Not of Hisp. Origin	96 (49)	99 (87)	94 (39)
White	99 (85)	100 (96)	99 (81)
Black	98 (67)	100 (93)	97 (57)
AI,Eskimo,Aleut	100 (99)	100 (100)	100 (99)
Asian,PI	100 (97)	100 (95)	100 (98)
Other race	100 (100)	100 (100)	100 (100)

Figures are rounded

In Oakland, the median percent contribution from the NRFU sampling error component to the total sampling error is over 94% for each of the redistricting items. However, at the tract level, the NRFU component contributes less to the total sampling error. In fact, for total population estimates at the tract level, the median percent contribution from the NRFU component is 41%. The next section explains the variance modeling techniques that were designed to measure the relationship between the relative variances and their associated estimated totals.

4.0 Generalized Variance Functions

Standard errors need to be published with each block estimate. Complexity arises when blocks are aggregated to form congressional districts because covariances between blocks would need to be provided. For instance, there are about 4,100 blocks in Oakland, so there would be over 8.4 million covariances. Therefore, alternative ways to publish standard errors were evaluated. The major focus of the discussion is on generalized variance functions (GVFs). A GVF is basically a regression model that estimates the relationship between the estimated relative variances and the estimated totals. It helps reduce publication costs and arguably has the appealing characteristic of smoothing the directly calculated variance estimates. GVFs also may be used to generate standard errors for large unknown geographic entities like congressional districts. The estimated GVF equations can be provided to the user who can then substitute in an estimate of interest to get the resulting relative variance and consequently the standard error. Some related references on applications of generalized variance methods will provide the reader with a wide array of examples. These include Dajani (1996), Bieler and Williams (1990), Judkins and Wright (1990), and Krenzke (1995). Wolter (1985) is a general reference. The following sections describe the comparisons made that helped develop the final variance model.

4.1 Comparison of Methods

Three main generalized variance modeling methods were evaluated that allow data users to calculate the standard errors. Method 1 involved calculating the median of the block-level relative variances, V^2 (see below) and the tract-level relative variances for each redistricting item. The estimate of the relative variance for redistricting data item K, at the geographic level I, is calculated as:

$$V^2_{KI} = \frac{var(\hat{B}_{KI})}{(\hat{B}_{KI} + GQ_{KI})^2}$$

where GQ_{KI} is the group quarters count for redistricting item K and geographic level I, and $var(\hat{B}_{KI})$ is obtained as described in section 3. This methodology follows closely with what is explained in Fan (1983) and Griffin, Navarro, and Bates (1991). Method 2 involved calculating a different generalized variance function (GVF) for each of the 39 redistricting items. This method creates 39 sets of

parameters for each census test site. Method 3 involved calculating one GVF for each census test site, which will be used for all redistricting items. A limitation of this Method 3 is that the two parameter GVF assumes a constant design effect (i.e., variance under a complex design divided by the variance under simple random sampling). Therefore, the model may not fit well to some of the redistricting data items.

The following describes GVF Methods 2 and 3 in more detail. The GVF to be used in this comparison is,

$$1) \quad V_x^2 = V_y^2 + b\left(\frac{1}{x} - \frac{1}{y}\right)$$

where, x is the estimated total, V_x^2 is the relative variance of x (i.e., RSE squared), y is the estimate of the site total population, V_y^2 is the relative variance of y, and b is the estimated regression parameter from the model. This can be rewritten as, $V_x^2 = a + b / x$, where $a = V_y^2 - b / y$. This model ensures positive variances and controls the variances so that $V_x^2 = V_y^2$ when x = y. Due to this restriction on the model, we will not get the 'best' fit to the data points, however, it is presumed that it will improve the fit to larger unknown geographic entities (i.e., congressional districts), which are arguably more important. For Methods 2 and 3, the following procudures were used. The following data points were discarded when they satisfied the following, 1) x = 0, 2) the estimate comes from an ICM block sample (i.e., $var_{NRFU}(x)=0$), 3) var(x) = 0. Ten iterations of weighted squares regression were done. For each iteration, the weights were adjusted as, $weight = 1 / (V_{predict}^2)^2$. Observations were investigated if the absolute value of their standardized residual was greater than 3.5. The primary model evaluation tool was the median absolute relative deviation (ARD). We used the adjusted-R^2 as a secondary evaluation tool to confirm the median ARD results. The model evaluation tools used were the adjusted-R^2 values and the median absolute relative deviation (ARD). An appealing feature of the median ARD is that it allows one to measure the fit to subgroups of the modeled data. The ARD is calculated by,

$$ARD = 100 \times \frac{\left| V_{predicted,x}^2 - V_{observed,x}^2 \right|}{V_{observed,x}^2}$$

Figure 1 is a plot of the relative variance and the estimated totals on the log scale for one of the 39 redistricting data items. Method 1, the dashed line, does not fit the data well since it does not use the relationship between the relative variances and the estimated totals. The plot shows similar curves resulting from Method 2, the dotted curve, and Method 3, the solid curve. The median ARD for Method 1 was 80, for Methods 2 and 3 the median ARD was 54 and 56, respectively. Methods 2 and 3 being close in fit was typical for most of the 39 redistricting items. However, because Method 3 fit one model to all 39 data items, there were a small number of redistricting data items for which the model was poor. However, we wanted to keep the modeling process simple, so it was decided to use Method 3.

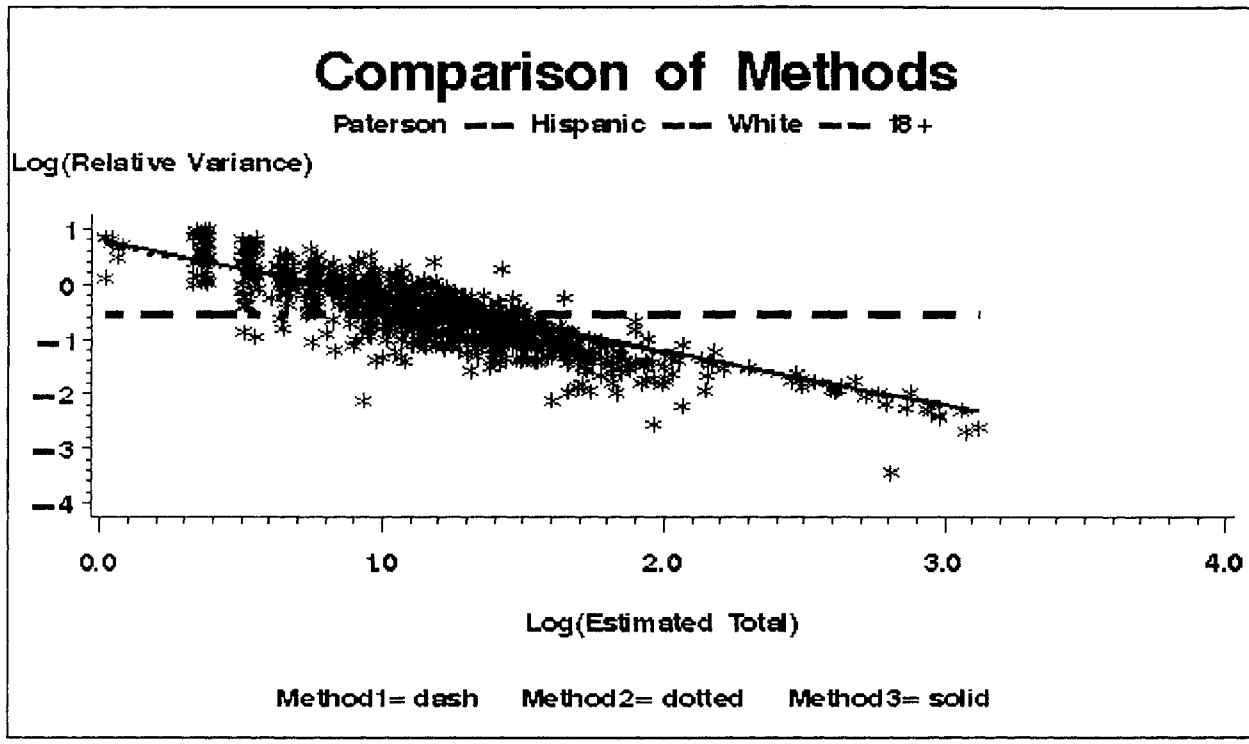
4.2 Comparison of Modeled Data

In a different comparison, the models were fit to block clusters and tracts separately, and to both block clusters and tract data together. In fitting the GVFs, the block clusters and tract data are completely dependent. However, combining the two geographic levels provides us with a compromise between the two separate models. More importantly, it provides a wide range of data with larger estimates which may yield more accurate variances for congressional districts. The results in Table 2 show no apparent difference between the 'b' parameters and median ARD, from the block cluster model and the block cluster/tract model. However, the block cluster/tract model is more appealing since it uses a full range of data. For instance, in Oakland, the largest block cluster size is about 1600, but the largest tract-level estimate is about 9600. Another point to make is that the tract model gives slightly higher standard errors since the 'b' parameters are larger. This may be due to the clustering of blocks within tracts which increases effect the variances. It was decided, however, that the block cluster/tract model should be used since it uses a wider range of data and uses two geographic levels. In the future, compromises such as using two or more geographic levels may be needed to approximate the variances for estimates the size of congressional districts. The highlighted parameters in Table 2 define the final GVFs.

Table 2. Comparison of Modeled Data

Site	Data	a	b	Med. ARD
Oakland	Block	.000362	2.008873	66.0
	Tract	.000360	2.815990	53.7
	Block/Tract	**.000362**	**2.061328**	65.4
Paterson	Block	.000316	5.784842	52.6
	Tract	.000299	8.352548	46.6
	Block/Tract	**.000315**	**5.922080**	51.9
NW La.	Block	.000193	2.229869	71.1
	Tract	.000169	5.110656	58.5
	Block/Tract	**.000193**	**2.292658**	71.4

Figure 1. Comparison of Methods



4.3 Comparison of Models

Several models were compared to Model 1 using the block cluster/tract data together. Most of the models were extensions of Model 1 as shown in Figure 2.

Figure 2. Comparison of Models

1) $V_x^2 = V_y^2 + b(\frac{1}{x} - \frac{1}{y})$

2) $V_x^2 = V_y^2 + b(\frac{1}{x} - \frac{1}{y}) + c(\frac{1}{\sqrt{x}} - \frac{1}{\sqrt{y}})$

3) $V_x^2 = V_y^2 + b(\frac{1}{x} - \frac{1}{y}) + c(\frac{1}{x^2} - \frac{1}{y^2})$

4) $V_x^2 = V_y^2 + b(\frac{1}{x} - \frac{1}{y}) + c(\frac{1}{\sqrt{x}} - \frac{1}{\sqrt{y}}) + d(\frac{1}{x^2} - \frac{1}{y^2})$

5) $V_x^2 = V_y^2 + b(\frac{1}{x} - \frac{1}{y}) + c(\frac{1}{\sqrt{x}} - \frac{1}{\sqrt{y}}) + d(\sqrt{x} - \sqrt{y})$

6) $V_x^2 = V_y^2 + b(\frac{1}{x} - \frac{1}{y}) + c(\frac{1}{x^2} - \frac{1}{y^2}) + d(x^2 - y^2)$

7) $V_x^2 = \frac{1}{(\frac{1}{V_y^2} + b(x-y))}$

The data showed no overwhelming support for any particular model. Model 1 compared favorably amongst the other models for each site and is also the simplest model of the seven that were compared. In addition, Model 1 is the only model having the desirable property where var(p) = var(1-p), where p is a proportion (Tomlin 1974). Therefore, Model 1 was the model chosen to present the standard errors to the data users. The final models were verified by block-group level variances. The median ARD was calculated to see how well the models fit the block-group data.

Standard errors for the estimated number of persons for a domain of interest may be calculated by using the following formula:

$$se(x) = \sqrt{ax^2 + bx}$$

where, x = estimated number of persons, and a and b are variance model parameters. The formula to estimate the standard error for a proportion of persons in a domain of interest, where p = x/y, is derived from an approximation to the Taylor Series formula (i.e., $V_p^2 \approx V_x^2 - V_y^2$) where y = estimated base population. Then the standard error for p:

$$se(p) = \sqrt{(\frac{b}{y})(p(1-p))}$$

5.0 Summary

Direct variances were calculated for 39 redistricting items for three 1995 Census Test sites. These variances include components from two sources of sampling error, error in estimation due to ICM sampling and error in estimation due to NRFU sampling. The statistical relationship between estimated totals and their associated directly calculated variance estimates was modeled for each site. Three ways to proceed with the modeling were compared, three data sets containing different combinations of geographic levels were used and the resulting model parameters were compared, and seven variance models were evaluated. The result of the modeling procedures is to use one GVF for each of the three Census Test sites to calculate the standard errors for estimated totals and proportions for all 39 redistricting items.

6.0 Future Research

This paper serves to document the beginning of research into ways of measuring the sampling error in Census 2000. It is the authors hope that this paper generates ideas on enhancing the methodology that was implemented in the 1995 Census Test and to generate ideas on alternative ways of measuring sampling errors. More research is needed to examine the effects of the assumptions on the resulting variance estimates. Also, the variance modeling procedure, if it is to be used in Census 2000, needs to be refined because the 1995 Census Test sites were well-defined areas that were smaller than congressional districts.

7.0 Acknowledgments

The authors acknowledge the work contributed by members of the Census Sampling and Estimation Staff at the Census Bureau, most notably, Henry Woltman for his guidance and valuable discussion and comments, Raj Singh for chairing the meetings and for his review and comments, William Bell and Cary Isaki for their work and comments on the ICM variance component, and Mary Mulry's review and discussion. All computer programs were written in the SAS programming language. Thanks to Jim Treat for supplying the data files. Thanks to Jenny Thompson for her valuable comments during the Census Bureau review.

[1] This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.

8.0. References

Bell, W. (1996), "An Approach to ICM Block Variances for the 1995 Test Censuses", Census Memorandum to Rajendra Singh dated February 12, 1996.

Bieler, G. And Williams, R. (1990), "Generalized Standard Error Models for Proportions in Complex Design Surveys," Proceedings of the Section on Survey Research Methods, American Statistical Association.

Dajani, A. (1996), "Generalized Variances for the 1990 SIPP Panel", Demographic Statistical Methods Division Report Series #96-01.

Fan, M. (1983), "Preliminary Summary of Results from a Comparison of Methods to Present 1980 Census Variance Estimates", 1980 Census Preliminary Evaluation Results Memorandum No. 54, 1983.

Griffin, R., Navarro, A., and Bates, L. (1991), "Generalized Variance Estimates Due to Adjustment of the 1990 Census", Proceedings of the Section on Survey Research Methods, American Statistical Association.

Judkins, D. And Wright, D. (1990), "National Health Interview Survey Report on Variance Estimation", WESTAT Series 2 Report dated September 26, 1990.

Krenzke, T. (1995), "Reevaluating Generalized Variance Model Parameters for the National Crime Victimization Survey", Proceedings of the Section on Survey Research Methods, American Statistical Association.

Killion, R. A. (1996), "The 1995 Census Test: A Compilation of Results and Decisions", DMD 1995 Census Test Results Memorandum No. 46, April 1, 1996

Schindler, E, and Navarro, A, (1994), "Census Plus: An Alternative Coverage Methodology", Proceedings of the Section on Survey Research Methods, American Statistical Association.

Tomlin, P. (1974), "Justification of the Functional Form of the GATT Curve and Uniqueness of Parameters for the Numerator and Denominator of Proportions" Unpublished memorandum, U.S. Bureau of the Census

Town, M., and Fay, R. (1995), "Properties of Variance Estimators for the 1995 Census Test", Proceedings of the Section on Survey Research Methods, American Statistical Association.

Wolter, K. (1985), *Introduction to Variance Estimation*, New York: Springer-Verlag.